EXHIBIT B

Page 1

IN THE UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

_____

RICHARD KADREY, et   al.,   )
)
Individual and Representative   ) Lead Case No.
Plaintiffs,   )
) 3:23-cv-03417-VC
    v.   )
)
Meta Platforms, Inc.,   )
)
Defendant.   )
_____)

** H I G H L Y   C O N F I D E N T I A L **

30(b)(1) VIDEOTAPED DEPOSITION OF

MICHAEL PATRICK CLARK

Denver, Colorado

Wednesday, November 13, 2024

Reported stenographically by:

Michelle Kirkpatrick, RDR-CRR-CRC-CRI, FCRR

_____

DIGITAL EVIDENCE GROUP

1730 M Street, NW, Suite 812

Washington, D.C. 20036

(202) 232-0646

Page 102

```
 1   something that would -- that there was code to                11:44

 2   basically -- you said remove, potentially, if it              11:44

 3   mentioned, you know, the source of the content?               11:44

 4          Is that something -- can you be more                   11:44

 5   specific on what that removal would look like?                11:44

 6          MS. HARTNETT:  Object to the form.                     11:45

 7      A   I'm a little confused by that question, so             11:45

 8   if you could provide a little more detail.                    11:45

 9   BY MR. STEIN:                                                 11:45

10      Q   Sure.  Yeah.                                           11:45

11          So I think you mentioned that if the data             11:45

12   was self-identified as pirated, that that could get          11:45

13   removed in the mitigation process.                           11:45

14          What exactly would be removed?                        11:45

15      A   So if the dataset was a series of files               11:45

16   and the file name said "pirated X, Y, Z" or "stolen          11:45

17   X, Y, Z," then the provenance of that data was very,         11:45

18   very clear, and that was pulled out of                       11:45

19   pre-processing as one of the early stages.                   11:45

20      Q   Okay.  And in terms of the -- you                     11:45

21   mentioned file name.                                         11:45

22          Is there anything in the content of, let's            11:45
```

Page 106

```
 1   removed, you're talking --                           11:48

 2      A    File.                                         11:48

 3      Q    -- about a file?                              11:48

 4      A    A file.                                       11:48

 5      Q    Okay.  So a file in a particular dataset      11:49

 6   might get removed because it was self-identified as   11:49

 7   pirated or stolen.                                    11:49

 8      A    Yes.                                          11:49

 9      Q    And my question is:  Why would that file      11:49

10   get removed?                                          11:49

11          MS. HARTNETT:  And I would just object to      11:49

12   the extent that there's any attorney-client-          11:49

13   privileged information that would be divulged by an   11:49

14   answer.                                               11:49

15      A    Yeah, that mitigation -- the reason for       11:49

16   that was a part of attorney-client with the IP and    11:49

17   product counsel teams.                                11:49

18   BY MR. STEIN:                                         11:49

19      Q    And were there -- you know, you mentioned     11:49

20   this example, I think you said was in a foreign       11:49

21   exchange.                                             11:49

22          Can you think of another example of an         11:49
```

Page 178

```
 1   it's -- if it self-identified and it is, then we        14:01

 2   know.  I think everything outside of that, we don't      14:01

 3   have a technical way of actually knowing where it        14:01

 4   originated and how that happened.                        14:02

 5           I also think that, you know, the whole           14:02

 6   process of training data is fully transformational       14:02

 7   and well within the boundaries of acting ethically       14:02

 8   and following the laws around fair use.                  14:02

 9           In addition to the mitigations that we did       14:02

10   on top of that to make sure that the model could not     14:02

11   regurgitate or memorize that material, I don't have      14:02

12   any reason to say it shouldn't.                          14:02

13   BY MR. STEIN:                                            14:02

14      Q    Why is it different from the data that you       14:02

15   did remove that was self-identified as pirated or        14:02

16   stolen?  Why should that matter?                         14:02

17      A    The source of the decision to do that            14:02

18   mitigation was direct advice from counsel.               14:02

19      Q    And your testimony is, even knowing that a       14:02

20   dataset contained pirated or stolen material, if it      14:03

21   had not self-identified as such, it does not need to     14:03

22   be removed.                                              14:03
```